DOCUMENT RESUME

ED 053 207 TM 000 739

AUTHOR	Evans, Franklin R.; Reilly, Richard
TITLE	A Study of Speededness as a Source of Test Bias.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	RB-71-31
PUB DATE	May 71
NOTE	18p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Fees, Financial Support, Item Analysis, *Negroes, *Reading Comprehension, Scores, *Test Bias, Testing Problems, *Test Reliability, Test Results, Tests, *Timed Tests
IDENTIFIERS	*Law School Admission Test, LSAT

ABSTRACT

Specially constructed "speeded" and "unspeeded" forms of a Reading Comprehension section of the Law School Admission Test (LSAT) were administered to regular center and fee-free center LSAT candidates in an effort to determine: (1) if the test was more speeded for fee-free candidates, and (2) if reducing the amount of speededness was more beneficial to fee-free candidates. Results of the analysis show: (1) the test is somewhat more speeded for fee-free candidates than for regular candidates, (2) reducing the amount of speededness produces higher scores for both regular (22 scaled score points) and fee-free (33 scaled score points) center candidates, and (3) reducing speededness is not more beneficial (in terms of increasing the number of items answered correctly) to fee-free than to regular center candidates. Lower KR-20 reliability was observed under speeded conditions in the fee-free sample and is discussed. (Author)

RESEARCH BULLETIN

ED053207

TM 000 739

RB-71-31

A STUDY OF SPEEDEDNESS AS A SOURCE OF TEST BIAS

Franklin R. Evans

and

Richard Reilly

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS [redacted] MATERIAL HAS BEEN GRANTED BY Educational Testing Service
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the authors. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
May 1971

A Study of Speededness as a Source of Test Bias

Franklin R. Evans and Richard Reilly

Abstract

Specially constructed "speeded" and "unspeeded" forms of a Reading Comprehension test were administered to regular center and fee-free center LSAT candidates in an effort to determine: (1) if the test was more speeded for fee-free candidates, and (2) if reducing the amount of speededness was more beneficial to fee-free candidates. Results of the analyses show: (1) the test is somewhat more speeded for fee-free candidates than for regular candidates, (2) reducing the amount of speededness produces higher scores for both regular (22 scaled score points) and fee-free (33 scaled score points) center candidates, and (3) reducing speededness is not more beneficial (in terms of increasing the number of items answered correctly) to fee-free than to regular center candidates.

Lower KR-20 reliability was observed under speeded conditions in the fee-free sample and is discussed.

A Study of Speededness as a Source

of Test Bias

Franklin R. Evans and Richard Reilly

It has often been charged that standardized tests, especially those used for academic admissions, are unfair (biased) to members of minority groups such as Blacks, Chicanos, and Puerto Ricans. Most of the recent studies of admissions test bias have dealt with the use of test scores as predictors of academic performance. Admissions tests are considered biased if predictions based on the test, either solely or in part, result in under-prediction for one particular group, e.g., Cleary (1968). This definition of bias, however, is dependent on some criterion which is itself assumed to be free of bias. It is this assumption of an unbiased criterion which may account for some of the conflicting results in test bias research. In academic admissions research the criterion against which test effectiveness is most often assessed is some form of earned grades. In the absence of specific knowledge about how much bias, if any, exists in the criterion, some criterion-free method may help to determine if a test is biased against one specific group. One reasonable approach is to (1) hypothesize what the specific biasing agent is (e.g., test length), (2) create a test which is relatively free of that biasing agent and (3) administer both the "biased" test and the "unbiased" test to the groups of interest and assess the effects.

The Law School Admission Test (LSAT) is designed to be a measure of accuracy and not speed, i.e., most candidates are expected to answer all of the questions on the test in the allotted time. It is probable, however, that Black candidates do not finish the test in as large a proportion as

White candidates do. To the extent the above is true there exists a potential source of bias against Blacks.

The present study was conducted in an effort to determine (1) if the Reading Comprehension section of the LSAT is more speeded for candidates from predominantly Black colleges than for a typical candidate population, and (2) if reducing the amount of speededness has a differential effect on the two candidate populations, i.e., is the result a larger score increment for one group.

Procedure

Subjects

For the past several years the Law School Admission Test Council (LSATC) has maintained special test centers on approximately 50 college campuses whose student populations are predominately Black. On one regular LSAT test date (usually February) each year the LSAT is provided free of charge to students on these campuses. The special center samples were comprised of all college senior candidates who took the LSAT in February 1970 at these fee-free centers. These students were divided into two groups: speeded (FFS), N = 230, and unspeeded (FFU), N = 235. The group a given student was assigned to was determined by which form of the LSAT he received when he reported to the test center. The test forms were arranged alternately so that every second subject received the same form. While the process is not truly random it is possible to assume that the two groups thus obtained were comparable.

Two samples were also drawn from all regular February 1970 LSAT centers in the continental United States. In the regular centers there were two forms

of the LSAT being used to pretest items so it was necessary to spiral the two research forms (speeded and unspeeded) with the two existing LSAT forms for these centers. Thus, every fourth student received the same form. There were approximately 23,000 students tested at all regular centers in February 1970. Of these approximately 5000 took the special speeded test and approximately 5000 the special unspeeded test. Spaced samples (1 of 4) of each of these populations yielded a regular center speeded sample (RCS) of 1250 and a regular center unspeeded sample (RCU) of 1245.

Experimental Subtests

In order to investigate the questions posed above it was necessary to create experimental forms of a test which differed only in the degree to which they were speeded. It is possible to do this in at least two ways: (1) administer identical tests under different time conditions and (2) administer tests with different numbers of items under the same time conditions. In this study the latter procedure was chosen in order that the research could be carried out in a regularly scheduled LSAT administration. (It was not possible to have different time limits for candidates who were being tested at the same location.) The "speeded" test was a 4-passage Reading Comprehension section and its related 35 items (each passage had 8 or 9 items related to it). The time limit for this 35-item section was 40 minutes. A 27-item "unspeeded" measure was created by eliminating the second passage and its 8 related items from the "speeded" measure. The time limit for this section was also 40 minutes. Thus, the two measures (speeded and unspeeded) had 27 items in common. All analyses pertaining to Section 6 were done on the scores derived from these 27 items. Even though the speeded

and unspeeded tests contained the same 27 items, they should not be considered the same test since the speeded version contained 8 additional items which could have affected the scores in some unknown way.

Results and Discussion

This study was designed in an attempt to answer the following questions. Is the Reading Comprehension section of the LSAT more speeded for test candidates who attend predominantly Black colleges than for a typical LSAT candidate population? What is the effect on the scores of the fee-free candidate as compared to the typical LSAT candidates of removing speededness? No attempt is or should be made to compare the LSAT scores of Black and White candidates. The focus of the study is on the differential effects of one potential source of bias, speededness.

The first question can be answered in a fairly straightforward fashion. According to criteria used by Swineford (1956) a test may be considered unspeeded if: (1) virtually all candidates reach 75 per cent of the items, and (2) at least 80 per cent of the candidates respond to the last item.

Insert Figures 1a and 1b about here

Figures 1a and 1b show the proportion of each of the four samples who reached each item of Sections 1 and 6 (Reading Comprehension) of the LSAT, respectively. The reader can see by looking at Figure 1a that Section 1 of the LSAT is clearly speeded for the fee-free samples under the criteria established above. Less than 85 per cent of the candidates in these centers reached item 19 while only about 60 per cent attempted the last item. For Section 6 two of the samples (speeded) were faced with the task of reading

four passages and answering 35 items in 40 minutes while the other two samples (unspeeded) were required to read only three passages and answer 27 items in the same amount of time. Since the samples from the fee-free centers are roughly comparable to each other and the samples from the regular centers are roughly comparable to each other, we can observe the effects of requiring fewer items on each of the two candidate populations. Section 6 is clearly a speeded measure for fee-free candidates when administered under the usual conditions. The fee-free candidates who had only 27 items to answer attempted considerably more of the last few items than those who had 35 items to answer. More than 90 per cent of the unspeeded group and less than 70 per cent of the speeded group attempted the last item. On the other hand, the elimination of 8 items did not appear to affect the regular center candidates nearly as much. More than 93 per cent of the unspeeded group as compared to more than 90 per cent of the speeded group attempted the last item.

Figures 1a and 1b clearly indicate that the Reading Comprehension section of the LSAT would be considered a speeded measure for fee-free candidates and an unspeeded measure for regular center candidates. Further, eliminating eight of the 35 items in Section 6 appears to remove much of the speed component for the fee-free candidates.

When interpreting these results one must remember that the LSAT score is based on the number of items answered correctly. No correction for guessing is applied and such is stated in the directions to the examinee. The fact that large numbers of candidates from the fee-free centers dropped out suggests that either the test was entirely too speeded for them or that the instruction to answer every item was not fully understood.

Insert Table 1 about here

Table 1 presents the means and standard deviations for LSAT total score and Section 1 score and Section 6 score for each of the four samples. The Section 6 score presented is a converted score based on the common 27 items of Section 6. These items were scored and the mean and standard deviation were then set equal to the mean and standard deviation of the LSAT for the sample of regular center candidates taking Section 6 under normal (speeded) conditions. Section 1 scores were derived in the same manner. Although the steps followed in this procedure were identical to those typically used in a linear equating of test scores, the converted Sections 1 and 6 scores should by no means be considered "equated" to the LSAT. A number of important assumptions necessary for a linear equating were not met, the most important of these being the assumption of parallel measures. The conversion was carried out simply to give the reader some idea of the magnitude of the results in terms of the usual LSAT scale. The scores of the fee-free groups can be seen to be much lower than the regular center groups on Sections 1 and 6. The fee-free unspeeded group scored approximately 33 scaled score points higher on Section 6 than the fee-free speeded groups while the difference between the Section 6 scores for the regular center groups was approximately 22 scaled score points.

In order to assess the reliability of the observed difference in Section 6 scores between fee-free centers and regular centers a 2 x 2 analysis of variance was performed. The dependent variable in this analysis was the number right on the 27 items of Section 6 that were common to both the speeded and unspeeded forms. In order to make the number of candidates in each group approximately equal every fifth regular center candidate was chosen for the analysis. The results of this analysis are shown in Table 2.

Insert Table 2 about here

The significant main effect for centers is of little interest since it merely confirms one of the results of Pitcher and Schrader (1968), namely that regular center candidates score higher on all sections of the LSAT than do fee-free candidates. Likewise, the main effect for Test Form simply shows that candidates, regardless of their test center category, score higher under unspeeded conditions. What is of most interest is the interaction between test center and test form. The absence of an interaction effect serves to reject the hypothesis that reducing the number of items in a given time period is of more benefit to fee-free candidates than regular center candidates. Thus, reducing the number of items in the Reading Comprehension section of the test will result in higher scores on those sections for all candidates, fee-free or regular center, but the amount of this increase could not be said to be significantly greater for free center candidates.

Table 3 shows the mean number of items not reached, the mean number of items omitted and the reliabilities for Sections 1 and 6 of the test for each of the four groups. Again we can see that the fee-free groups apparently reached fewer items than the regular center groups in Section 1 and the unspeeded Section 6. They also appear to have omitted a somewhat larger number of items than the regular center groups. If all of the fee-free candidates had guessed on the items not reached or omitted in Section 1, they would be expected to increase their raw score mean by approximately .57. Likewise, the regular center groups would be expected to gain approximately

.12 raw score points on their Section 1 mean. This pattern is similar for the speeded groups on Section 6 with the expected gain from guessing being .65 and .10 for the fee-free and regular groups, respectively. On the other hand, the gain expected in the unspeeded samples would be .20 and .03 for the fee-free and regular center groups, respectively. With these results one could conclude that chance alone would account for about .45 raw score points or approximately 1/3 of the raw score difference that was observed between the fee-free speeded and unspeeded Section 6 scores.

Ordinarily, when an unspeeded test is made speeded the expectation is that odd-even (split-half) test reliability will appear higher because lower ability candidates drop out before reaching the last few items. The effect of speededness on Kuder-Richardson reliabilities has not been clearly established. However, since KR-20 is a function of item variances and covariances we would expect speededness to have some effect. For the regular center candidates little difference would be expected since the test did not appear to be speeded under either condition. The KR-20 reliabilities of .71 under unspeeded conditions and .72 under speeded conditions confirmed this expectation. For the fee-free candidates, however, the unspeeded reliability of .73 was significantly ($p < .05$) larger than the speeded reliability of .59. (This test was made using a procedure for establishing confidence limits on KR-20 reliabilities due to Feldt (1965).) This finding may be related to differences in the Section 6 standard deviations (see Table 1) of the two fee-free groups (speeded vs. unspeeded).

Insert Table 3 about here

A test for homogeneity of variance (Winer, 1962, p. 34) revealed that the variance observed in the unspeeded sample was larger than that observed in the speeded sample ($p < .02$). Thus, decreasing the number of items in the fee-free centers increased the variability of the Section 6 scores which in turn appears to have had the effect of making those scores more reliable.

As discussed earlier, virtually all of the regular center candidates attempted the last few items of Section 6 under the speeded condition. Since there was no penality for guessing this was to be expected. The candidates from the free centers, on the other hand, dropped out in large numbers near the end of the test under the speeded condition. Whatever the cause we must consider that requiring the fee-free group to take the test under highly speeded conditions results in less variable and less reliable scores for them. Since it is likely that the entire LSAT is somewhat more speeded for the fee-free group the result could be that we are also obtaining less reliable LSAT scores for them.

Summary and Conclusions

Specially constructed "speeded" and "unspeeded" forms of a Reading Comprehension test were administered to regular center and fee-free center LSAT candidates in an effort to determine: (1) if the test was more speeded for fee-free candidates, and (2) if reducing the amount of speededness was more beneficial to fee-free candidates. Results of the analyses show: (1) the test is somewhat more speeded for fee-free candidates than for regular candidates, (2) reducing the amount of speededness produces higher scores for both regular (22 scaled score points) and fee-free (33 scaled score points) center candidates, and (3) reducing speededness is <u>not</u> more beneficial

(in terms of increasing the number of items answered correctly) to fee-free than to regular center candidates.

Lower KR-20 reliability was observed under speeded conditions in the fee-free sample and is discussed.

References

Cleary, T. A. Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Feldt, L. The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. Psychometrika, 1965, 30(3), 357-370.

Pitcher, B., & Schrader, W. B. Item and test analysis results for candidates tested at the free LSAT administration of February 1967. Law School Admission Test Council Annual Report - 140-151 (1968). Princeton, N. J.: Educational Testing Service, 1968.

Swineford, F. Technical manual for users of test analyses. Statistical Report 56-42. Princeton, N. J.: Educational Testing Service, 1956.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Table 1

Means and Standard Deviations for LSAT Total,

Section 1 and Section 6

Samples	LSAT Total		Section 1*		Section 6**	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Fee-free speeded (FFS)	341.19	85.53	390.06	78.70	381.87	82.17
Fee-free unspeeded (FFU)	343.50	93.11	394.10	81.30	415.13	101.35
Regular speeded (RCS)	514.08	103.44	514.08	103.44	514.08	103.44
Regular unspeeded (RCU)	512.45	99.05	513.03	99.87	536.20	99.32

*Estimated scores $X' = ax + b$ where
$a = 22.79$
$b = 166.93$

**Estimated scores $X' = ax + b$ where
$a = 22.73$
$b = 214.10$

Table 2

Source	df	Mean Square	F
Centers	1	6708.98	347.99*
Test Form	1	386.29	20.04*
Centers x Form	1	10.66	.55
Error	960	19.28	
Total	964		

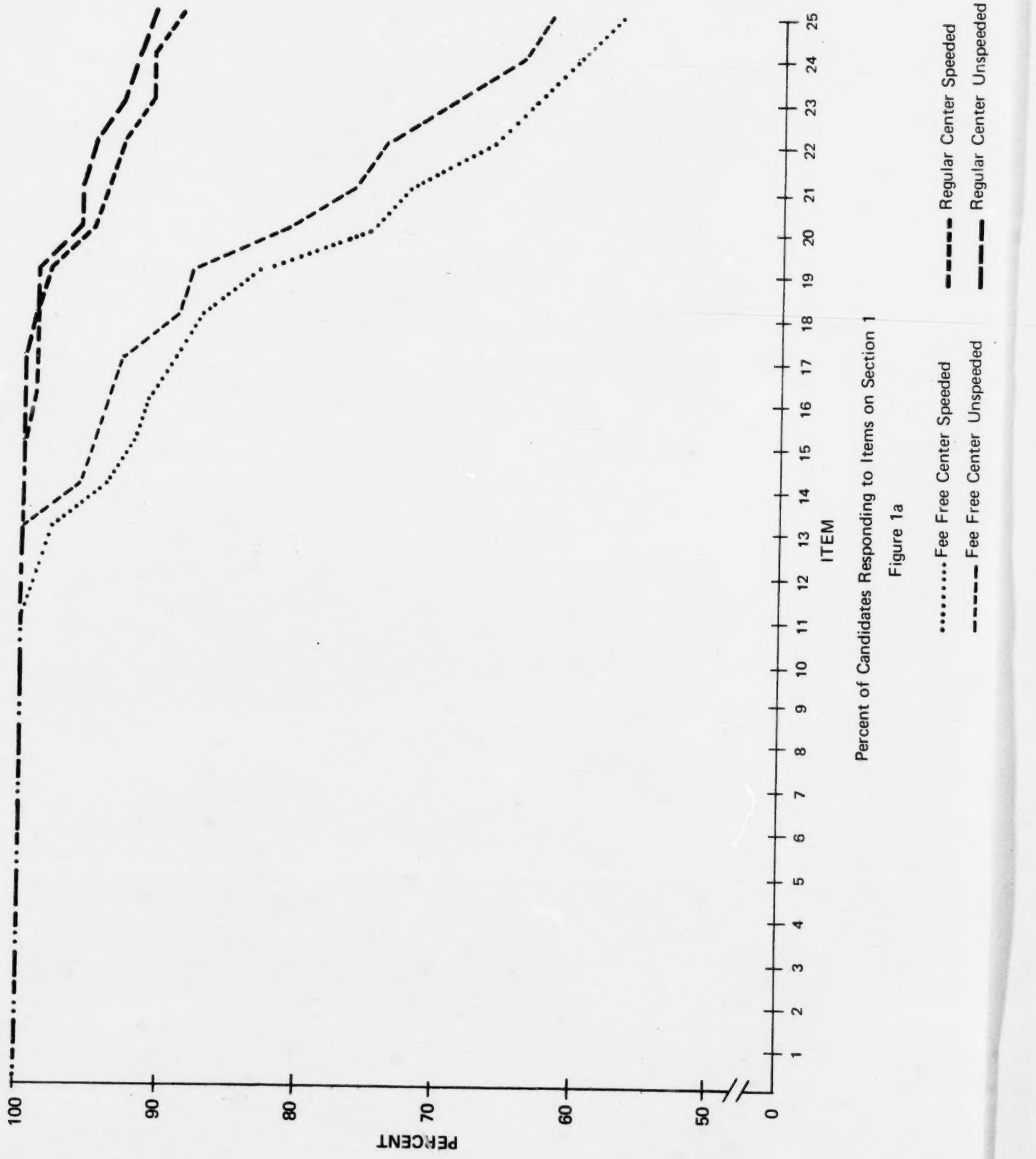
*p < .01

Table 3

Mean Number of Items Omitted and Not Reached and KR-20

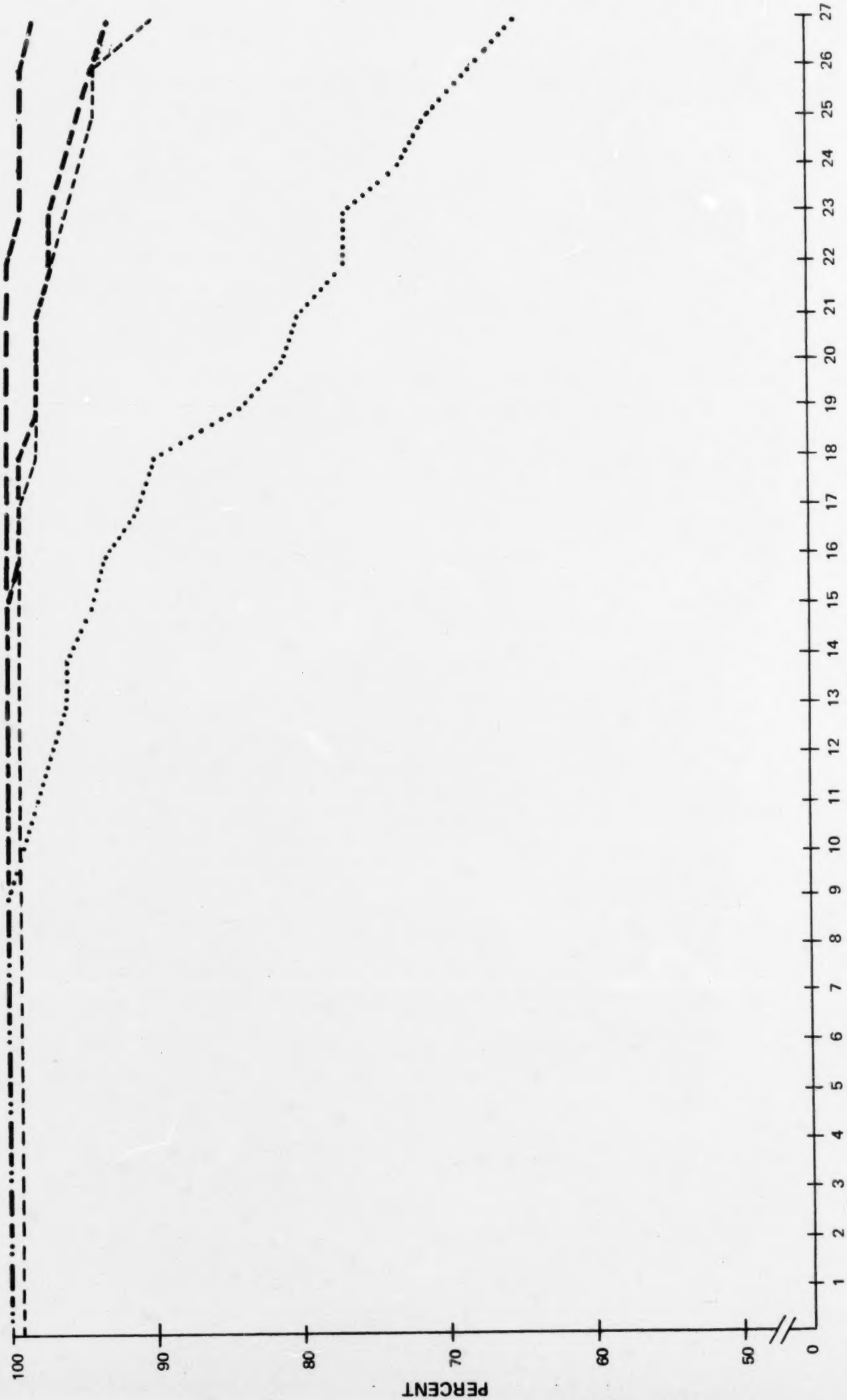
Reliabilities for Sections 1 and 6

	Section 1				Section 6			
	FFS	FFU	RCS	RCU	FFS	FFU	RCS	RCU
Items	25	25	25	25	27(35)	27	27(35)	27
N	230	235	1250	1245	230	235	1250	1245
Mean Not Reached	2.74	2.18	.53	.42	2.73	.59	.38	.11
Mean Omitted	.32	.46	.1328	.1016	.54	.36	.14	.05
Reliability	.61	.63	.75	.73	.59(.62)	.73	.72(.76)	.71



Percent of Candidates Responding to Items on Section 1

Figure 1a



Percent of Candidates Responding to Items on Section 6

Figure 1b